

Arto, Fennica ja Linda kansallisen julkaisurekisterin tietolähteinä

Tiivistelmä:

Kansalliskirjaston ylläpitämiä viitetietokantoja (Arto ja Fennica/Linda) on mahdollista käyttää kansallisen julkaisurekisterin kotimaisia julkaisuja koskevien tietojen lähteenä. Lyhyellä tähtäimellä (Jure I) hyödyt ovat tosin rajallisia, sillä tietokannat eivät tällä hetkellä sisällä mm. affiliaatitietoja, mikä käytännössä tarkoittaa sitä, että tietoja voi haravoida ainoastaan yksittäin tekijän tai artikkelin nimen perusteella, ei keskitetysti organisaatio kerrallaan affiliaatioiden perusteella.

Pidemmällä tähtäimellä (Jure II) kansallisia tietokantoja ja niiden tietosisältöjä on kuitenkin mahdollista kehittää siten, että ne palvelevat nykyistä paremmin myös julkaisurekisterien tarpeita. Etenkin Arton kehittäminen avoimeksi artikkeliviitteitä sisältäväksi tietovarannoksi olisi monin tavoin hyödyllistä. Tallennettavien tietojen ja tallennuksen prosessien kehittäminen edellyttää nykyisten tiedontuottajien lisämotivointia ja/tai nykyistä läheisempää yhteistyötä kotimaisten tiedekustantajien kanssa. Kustantaja-yhteistyö ja sitä varten tarvittava uusi syöttöjärjestelmä avaisivat mahdollisuuksia sellaisen julkaisutietojen keräämiseen ja hallintaan soveltuvan kansallisen kokonaisjärjestelmän rakentamiseen, joka palvelisi julkaisurekisterin lisäksi myös tiedonhakijoiden ja tieteellisten kustantajien tarpeita.

1. Mistä kotimaisten julkaisujen tiedot?

Yksi kansallisen julkaisurekisterin suunnittelun lähtökohdista on ollut tiedontuotannon tehostaminen ja tietojen laadun parantaminen hyödyntämällä muista tietokannoista saatavia valmiita viitetietoja. Merkittävä osa suomalaisten tutkijoiden kansainvälisissä lehdissä julkaisemista artikkeleista löytyy Web of Science ja Scopus –tietokannoista, joista niitä koskevia tietoja on mahdollista haravoida lisenssisopimusten sallimassa laajuudessa sekä kunkin tutkimusorganisaation oman että yhteisen kansallisen julkaisurekisterin pohjaksi. Näiden tietojen avulla voidaan vähentää sekä paikallisesti tehtävää luettelointityötä että ehkäistä monen organisaation yhteisjulkaisuista aiheutuvia duplikaattiongelmia.

Kansainväliset tietokannat eivät kuitenkaan sisällä tietoja tutkijoiden kansallisilla foorumeilla ilmestyneistä julkaisuista. Niillä aloilla joilla julkaisutoiminta on pääosin kansainvälistä, tällä ei ole kovin suurta merkitystä, mutta etenkin humanistis-yhteiskuntatieteellisillä tieteenaloilla kotimaisella julkaisemisella on edelleen suuri rooli ja huomattava osa OKM:lle vuosittain raportoitavista julkaisuista ilmestyy nimenomaan kotimaisten julkaisukanavien kautta.

Tässä selvityksessä tarkastellaan sitä, miten Kansalliskirjaston ylläpitämiä kansallisia viitetietokantoja (Arto, Fennica ja Linda) olisi mahdollista hyödyntää Kansallisen julkaisurekisterin tietolähteinä ja millaisia kehitystarpeita niiden hyödyntämiseen liittyy. Hyvä vertailukohta löytyy monissa yhteyksissä muutenkin esimerkiksi käytetystä Norjasta, jossa jopa neljäsosa kansallisen julkaisurekisterin sisältämistä artikkelien viitetiedoista saadaan kansallisesta viitetietokannasta. Norjan opetusministeriö on myös tällä perusteella myöntänyt lisärahoitusta tietokannan ylläpitoon.

2. Arto, Fennica ja Linda

2.1. Arto – kotimainen artikkeliviitetietokanta

Kansalliskirjaston ylläpitämän Arto-tietokannan varhaisin edeltäjä oli Turun yliopiston kirjaston 1950-luvun lopulta lähtien julkaisema painetussa muodossa ilmestynyt artikkeli-indeksi, joka korvattiin 1980-luvun alussa KATI-artikkelitietokannalla. Turhan päällekkäisen työn välttämiseksi yhteiseen kansalliseen tietokantaan pyrittiin integroimaan myös eri organisaatioiden ylläpitämiä tieteenalakohtaisia artikkeli-indeksejä. Vuonna 1995 artikkelitiedon tallennus organisoitiin uudelleen siten, että tiedot tallennettiin silloisen Helsingin yliopiston kirjaston ylläpitämään Arto-tietokantaan. Nykyään Arto on Voyager-tietokanta, joka on käytettävissä verkon kautta osoitteessa <https://arto.linneanet.fi>.

Arton perusidea on ollut alusta lähtien se, että tietojen tallennus on hajautettu eri puolille kirjastoverkkoa siten, että kukin luettelointiyhteistyöhön osallistuva kirjasto vastaa tiettyjen lehtien luetteloinnista. Yhteistyöhön osallistuville kirjastoille ei ole maksettu rahallista korvausta, mutta ne ovat saaneet vastineeksi tietokannan käyttöoikeuden. Kyse on siis ollut hyvin pitkälle vapaaehtoisuuteen perustuvasta yhteisen hyvän tuottamisesta. Tietokannan tuottamat tulot on käytetty sen ylläpitoon (ml. palvelin- ja lisenssikulut sekä jonkin verran henkilötöitä).

Tällä hetkellä Arto-tallennukseen osallistuu 13 yliopistoa, 10 erikoiskirjastoa, 4 maakuntakirjastoa ja yksi ammattikorkeakoulu. Tietokanta sisältää noin 1,5 miljoonaa tietuetta, ja sen vuosikartunta on keskimäärin 40.000 artikkeliviitettä, joita esim. vuonna 2010 tallennettiin kattavasti yli 300 eri lehdestä. Lehtivalikoiman ytimen muodostavat tieteelliset lehdet, minkä lisäksi siihen kuuluu myös erilaisia ammattilehtiä, populaarimpia tiedelehtiä ja monenlaisia muitakin julkaisuja. Tieteellisten lehtien osalta tietokannan kattavuus on parhaimmillaan etenkin humanistis-yhteiskuntatieteellisillä aloilla, mutta se sisältää runsaasti myös muiden tieteenalojen julkaisuja (esim. maa- ja metsätieteet).

Arto-tietokannan tulevaisuuden kannalta merkittävin epävarmuustekijä on ollut kirjastojen sitoutuminen sen tuottamiseen. Arto-tallennuksen kaltaista yhteisen hyvän tuottamista ei ole yleensä laskettu kirjastojen ydintoiminnaksi, ja budjettien kiristyessä niillä on ollut houkutus keskittää resurssejaan yhteisen kansallisen tietokannan tuottamisen sijasta omien paikallisten palvelujensa kehittämiseen. Tämä on johtanut siihen, että Arton lehtilistasta on ajoittain poistunut julkaisuja, tai niitä luetteloimaan on jouduttu etsimään uusia vastuukirjastoja. Toisaalta tietokannan käyttö on kuitenkin kasvanut vuosi vuodelta, sillä useimpien siihen luetteloitavien tieteellisten lehtien kohdalla ainoa vaihtoehto Arton tarjoamalle tiedonhakumahdollisuudelle on alkuperäisten painettujen julkaisujen selaaminen.

Vuoden 2010 alusta lähtien Arto on ollut OKM:n rahoituksella vapaassa käytössä. Ainakin toistaiseksi vapaan käytön rahoituksesta neuvotellaan Kansalliskirjaston ja OKM:n kesken vuosi kerrallaan. Arto-tietokannan vapaa käyttöoikeus avaa uusia mahdollisuuksia tietojen hyödyntämiseen ja tiedontuottajien piirin laajentamiseen.

2.2. Fennica ja Linda: vaihtoehtoja

Tieteellisiä monografioita koskevia tietoja on mahdollista haravoida julkaisurekistereihin joko Fennica- tai Linda-tietokannasta. Fennica (<https://fennica.linneanet.fi>) on Suomen kansallisbibliografia, ja se sisältää tiedot kaikista Suomessa vuosittain julkaistavista kirjoista. Linda (<https://linda.linneanet.fi>) taas on yliopistokirjastojen yhteistietokanta. Fennica on Voyager-tietokanta, kun taas Lindassa on siirrytty yhteistietokantaan liittyvistä teknisistä syistä johtuen käyttämään saman järjestelmätoimittajan toista kirjastojärjestelmäohjelmistoa, Alephia.

Fennican luettelointi perustuu periaatteessa kirjapainoilta saataviin vapaakappaleisiin. Koska vapaakappaleita toimitetaan monissa tapauksissa merkittäväällä viiveellä (esim. neljännesvuosittain tai jopa vuosittain), prosessia on nopeutettu tietuevaihdolla kirjastoalan kaupallisiin toimijoihin kuuluvan BTJ:n kanssa. Tämän tietuevaihdon ansiosta esim. kotimaisen kaunokirjallisuuden viitetiedot saadaan tietokantaan yleensä nopeasti. Tieteellisen kirjallisuuden osalta tilanne on usein kuitenkin huonompi.

Monista julkaisuista saadaan lisäksi tietokantaan suppeat ennakkotiedot Suomen ISBN-keskuksen kautta siinä yhteydessä kun julkaisija hankkii julkaisulleen ISBN-numeron. Suuremmat kustantajat jakavat kuitenkin ISBN-numeroita itse, eli niiden osalta julkaisujen tiedot eivät tule tätä kautta, ja lisäksi ennakkotiedot ovat julkaisurekisterin tarpeiden kannalta liian suppeita ja usein myös epämääräisiä. Kansalliskirjasto onkin pyrkinyt siihen, että julkaisujen tiedot saataisiin tulevaisuudessa suoraan kustantajien ja aineiston välittämien tietojärjestelmistä näiden käyttämässä Onix-formaatissa, josta se voidaan sitten konvertoida kirjastotietokantojen käyttämään Marc-formaattiin. Tämä tiedonsiirtomahdollisuus koskee kuitenkin lähitulevaisuudessa todennäköisesti pääasiassa suurten kaupallisten kustantajien julkaisuja, joiden tiedot on mahdollista saada valmiina muista tietojärjestelmistä.

Fennican keskeinen heikkous julkaisurekisterien tietolähteenä on siis tieteellisten julkaisujen tietojen suhteellisen hidas tallennusprosessi. Julkaisurekisterin kannalta on oleellista, että kunkin julkaisun tiedot ovat saatavilla mahdollisimman pian sen ilmestymisen jälkeen, jotta niitä voidaan hyödyntää vuosittaisessa raportoinnissa. Yliopistokirjastojen yhteistietokanta Linda saattaakin olla tältä osin parempi vaihtoehto, sillä uudet tieteelliset julkaisut luetteloidaan sinne usein nopeammin kuin Fennicaan.

Arton tavoin myös Linda on ollut vapaassa käytössä yliopistojen ulkopuolella vasta vuoden 2010 alusta lähtien, ja sen käyttöoikeudesta sovitaan vuosi kerrallaan. Yliopistokirjastot käyttävät nykyään uusien aineistojen luettelointiin nimenomaan Lindaa, josta julkaisujen tiedot siirretään sitten paikallisiin kirjastotietokantoihin, eli se tarjoaa tältäkin osin Fennicaa parempia edellytyksiä yliopistojen tekemän tiedontuotannon rationalisointiin ja nimenomaan julkaisurekisterien tarpeita palvelevien tietosisältöjen kehittämiseen. Lisäksi Lindan etuihin kuuluu myös se, että se sisältää kotimaisten julkaisujen lisäksi tietoja suomalaisten tutkijoiden ulkomailla julkaisemista monografioista ja kokoomateoksista, joista kaikki eivät välttämättä löydy kansainvälisistä tietokannoista.

Julkaisurekistereihin liittyvien prosessien suunnittelussa kannattaa huomioida myös suunnitteilla oleva Kansallinen yhteisluettelo. Tavoitteena on se, että tulevaisuudessa kaikki suomalaiset kirjastot luettelisivat aineistonsa samaan tietokantaan, joka toimisi samantyyppisen mallin mukaisesti kuin nykyinen Linda.¹ Yhteisluettelon lisäksi rakenteilla on myös kansallinen auktoriteettitietokanta, joka helpottaa mm. tekijätietojen hallintaa ja saman tekijän eri julkaisujen yhdistämistä toisiinsa.² Auktoriteettitietokannan ja OKM:n suunnitelmissa esillä olleen kansallisen tutkijarekisterin välille olisi luontevaa rakentaa kytkentöjä, joista olisi todennäköisesti merkittävää etua molemmille.

3. Kansalliset tietokannat julkaisurekisterien tietolähteinä

3.1. Nykytilanne

Teknisesti viitetietojen haravointi Artosta, Fennicasta tai Lindasta on mahdollista jo nyt, eikä tietojen käytölle yliopistojen julkaisurekistereissä ole käytännössä sopimuksellisiaakaan esteitä. Ongelmana on vain se,

¹ Nina Hyvönen & Petri Tonteri: Kansallinen yhteisluettelo -hanke. Tietolinja 2/2009. Luettavissa osoitteessa <http://urn.fi/URN:NBN:fi-fe200912102410>.

² Laila Heinemann: Asteri Lindalle: auktoriteettitietokantahankkeesta. Tietolinja 1/2010. Luettavissa osoitteessa <http://urn.fi/URN:NBN:fi-fe201005211889>.

että yliopistojen nykyiset julkaisurekisteri- tai tutkimuksenhallintaohjelmistot eivät paria poikkeusta lukuun ottamatta tue valmiiden viitetietojen haravointia sen paremmin koti- kuin ulkomaisistakaan tietokannoista.

Merkittävin poikkeus on Helsingin yliopiston vuonna 2010 käyttöön otettu tutkimuksenhallintajärjestelmä Tuhat (<https://tuhat.halvi.helsinki.fi/portal/>), jonne on mahdollista haravoida julkaisujen viitetietoja mm. Artosta ja Fennicasta. Haravointi toimii siten, että julkaisutietojaan tallentava tutkija hakee julkaisuja tietokannoista esim. omalla nimellään, tunnistaa tietokannasta löytyneistä tietueista omat julkaisunsa ja tallentaa ne sitten julkaisurekisteriin. Tuhat-järjestelmä huolehtii tuplakontrollista, eli se tarkistaa julkaisun tietoja analysoimalla onko se tallennettu julkaisurekisteriin jo aiemmin.

Vaikka Arto, Fennica ja Linda soveltuvat myös julkaisurekisterin tietolähteiksi, on hyvä huomata, että ne on alun perin kehitetty muuhun tarkoitukseen. Ne ovat ensisijaisesti tiedonhakijoita varten luotuja palveluita, minkä lisäksi Linda ja Fennica palvelevat kirjastojen prosesseja ja Fennica kansallisbibliografiana myös kirjalaa. Tietokantojen oletettu käyttötarkoitus näkyy myös niiden tietosisällöissä: tiedonhaun tarpeiden vuoksi julkaisujen asiansanoittamiseen ja muuhun sisällönkuvailuun on usein käytetty paljon resursseja. Julkaisurekisterin kannalta asiansanat eivät kuitenkaan ole lainkaan välttämättömiä, kun taas esim. tekijöiden affiliaatitiedot, julkaisutyyppien tarkempi määrittely tai julkaisun tieteenalan määrittäminen (etenkin jos se olisi tehty OKM:n tiedonkeruussa käytetyn luokituksen mukaan) olisivat paljon tärkeämpiä.

Käytännössä tekijöiden affiliaatitietojen puuttuminen on merkittävää, sillä se rajoittaa oleellisesti tietueiden haravointimahdollisuuksia. Jos nämä tiedot olisi tallennettu tietokantaan (joko henkilöihin yhdistettynä tai edes julkaisun tasolla), haravointia voisi tehdä myös niiden pohjalta keskitetysti siten, että tietokannasta haravoitaisiin kerralla koko organisaation julkaisut. Nykytilanteessa haravointia on mahdollista tehdä ainoastaan tekijän tai julkaisun nimen perusteella, jolloin julkaisun yhdistäminen johonkin tiettyyn organisaatioon tapahtuu vasta julkaisurekisterissä.

Muita julkaisurekisterien tarpeiden kannalta ongelmallisia piirteitä liittyy tietokantojen kattavuuteen ja tiedon ajantasaisuuteen. Esim. Arto-tietokanta ei kata tasaisesti kaikkia aloja, sillä siitä puuttuu sellaisia keskeisiäkin lehtiä joille ei ole löytynyt luetteloijaa. Lisäksi luetteloinnissa saattaa muutenkin olla aukkoja. Osittain tietokannan kattavuuden rajallisuutta selittää se, että joidenkin alojen (tekniikka, kauppatieteet, lääketiede, populaarijulkaisut) aineistoja luetteloidaan muihin kotimaisiin tietokantoihin, joita ovat mm. Aleksi, Tali, Finp ja Medic (muista kotimaisista artikkelitietokannoista ja niiden hyödyntämiseen liittyvistä kysymyksistä ks. tarkemmin LIITE 1). Lisäksi etenkin artikkelikokoelmien sisällön luettelointi Artoon on ollut perinteisesti puutteellista, sillä satunnaisemmin ilmestyvinä niitä koskevaa tallennusvastuuta on ollut vaikeampi jakaa tiedontuottajakirjastojen kesken.

Tietojen ajantasaisuuteen vaikuttavat monet tekijät. Fennica-tietokannan tallennuksessa julkaisuja luetteloidaan vapaakappaleaineistojen pohjalta, mikä aiheuttaa merkittäviä viiveitä. Lisäksi etenkin Arto-tallennusta tehdään kirjastoissa yleisesti muiden tehtävien ohessa, jolloin se vähemmän kiireelliseksi koettuna jää helposti odottamaan vuoroaan pitkäksiin aikaa. Julkaisurekisterin kannalta julkaisutietojen saaminen nopeasti olisi kuitenkin kriittistä, sillä raportointiaikataulu on etenkin loppuvuodesta tiukka, ja jos julkaisun kuvailutiedot eivät raportointivaiheessa ole saatavilla valmiina, ne täytyy odottelun sijasta joka tapauksessa tallentaa uudestaan.

3.2. Tietosisältöjen kehittäminen

Pidemmällä tähtäimellä etenkin Artoa on mahdollista kehittää siten, että se palvelee tiedonhakijoiden tarpeiden lisäksi nykyistä paremmin myös julkaisurekisterien ja tutkimushallinnon tarpeita. Jos Arto saadaan pysyvästi vapaaseen käyttöön, sitä voitaisiin kehittää avoimeksi kotimaisen artikkelitiedon tietovarannoksi, joka hyödyttäisi sekä tiedonhakijoita, tutkimushallintoa ja lehtien kustantajia. Lisäksi sen sisältämiä tietoja olisi mahdollista hyödyntää mm. julkaisuarkistoissa, artikkeliaineistojen digitoinnissa ja erilaisten muiden-

kin verkkopalveluiden rakentamisessa. Yhteisen avoimen tietovarannon kehittäminen liittyy kirjastojen keskuudessa yleistyvään Open Linked Data -ideologiaan,³ ja se tukee myös julkisilla varoilla tuotettujen tietovarantojen nykyistä avoimempaa käyttöä koskevien kansallisten linjausten⁴ toteutumista.

Affiliaatitietojen lisääminen Artoon ja muihinkin kansallisiin viitetietokantoihin on mahdollista, eikä siihen liity kovin merkittäviä teknisiä ongelmia. Kirjastotietokannoissa käytetyssä Marc-formaatissa on valmiina soveltuva alakenttä, jota voidaan käyttää tietojen tallentamiseen. Kuten jo aiemmin on tuotu esille, teki-
jöiden affiliaatitiedot saattaa olla mahdollista yhdistää myös tulevan kansallisen auktoriteettitietokannan sisältämiin tietoihin, eli niitä ei tarvitsisi tallentaa yksittäisiin tietueisiin. Koska tieteellisten julkaisujen ilmestymisessä on toisaalta usein kuukausien tai jopa vuosien viiveitä ja julkaisun tekijä on tällä välin saattanut siirtyä organisaatiosta toiseen, tämä ei kuitenkaan ole aivan yksinkertaista.

Käytännössä affiliaatitietojen tallentaminen aiheuttaa luetteloijille jonkin verran lisätyötä, eli aineistojen luettelointi hidastuu hieman. Kansainvälisiin julkaisuihin verrattuna kotimaisten julkaisujen kirjoittajamäärä on yleensä suhteellisen pieni, eli niissä harvemmin törmätään tilanteisiin, jossa yhdellä artikkelilla on kymmeniä tai jopa satoja kirjoittajia. Toisaalta kaikissa kotimaisissa tieteellisissä julkaisuissa ei ilmoiteta teki-
jöiden affiliaatioita tai ne ovat puutteellisia. Jos affiliaatioita ei ole lehdessä mainittu, luetteloijalla on hyvin rajalliset mahdollisuudet selvittää asiaa, tai ainakin selvitystyö on työlästä ja tulokset epävarmoja.

Affiliaatitietojen lisäksi Artoon olisi mahdollista tallentaa myös julkaisutyyppitieto (referee-artikkeli, kirja-arvio, jne), jos tieto on saatavissa julkaisusta. Julkaisuerekistereiden kannalta olisi hyödyllistä pyrkiä erottelemaan selkeästi toisistaan vertaisarvioitu ja vertaisarvioimaton materiaali, ja muutenkin päästä siihen, että lehden sisältö luokiteltaisiin siten, että sen ja lehden tieteellisyyttä koskevan tiedon (tieteellinen/ei-tieteellinen julkaisufoorumi) pohjalta olisi mahdollista johtaa automaattisesti myös OKM:n julkaisutyyppi-
luokituksen mukaiset julkaisutypit. Valitettavasti kaikki lehdet eivät ilmaise selkeästi, mitkä artikkelit ovat vertaisarvioituja ja mitkä eivät, jolloin luetteloijalla ei ole keinoja päätellä oikeaa julkaisutyyppiä.

Ainakin affiliaatitietojen ja referee-käytännön osalta olisi hyödyllistä, jos lehtien kustantajia voitaisiin kannustaa liittämään nämä tiedot selkeästi ja standardoidussa muodossa itse julkaisuun. Affiliaatitietojen osalta tämä edellyttää sitä, että tiedot kerätään tekijöiltä itseltään, eli tähän liittyvät ohjeet olisi hyvä saada mukaan kunkin lehden kirjoittajan ohjeisiin. Referee-käytäntöjen noudattamisen osalta paras tieto on puolestaan lehden toimituskunnalla. Tiedekustantajien käytäntöihin vaikuttaminen ei todennäköisesti onnistu aivan kädenkäänteessä, mutta asiaa olisi kuitenkin perusteltua ryhtyä edistämään esim. yhteistyössä Tieteellisten seurain valtuuskunnan ja Suomen tiedekustantajien liiton kanssa.

Viittaustietojen kokoaminen kansallisiin tietokantoihin Web of Sciencen tai Scopusen mallin mukaan ei toisaalta näytä ainakaan tällä hetkellä realistiselta vaihtoehdolta. Viittaustietojen tallentaminen vaatisi runsaasti työtä sekä tietojen tallentamiseksi että järjestelmien kehittämiseksi, eikä siitä todennäköisesti saataisi vastaavaa hyötyä, sillä järjestelmä voisi tunnistaa ainoastaan sellaiset viitattut teokset, jotka ovat mukana samassa aineistossa. Suuri osa kansallisten julkaisujen viittauksista kohdistuu kuitenkin kansainvälisiin tai muihin aineiston ulkopuolelle jääviin julkaisuihin, eikä järjestelmä tavoittaisi myöskään aineiston ulkopuolelta tulevia viittauksia.

³ Laila Heinemann: Kirjastodataa raakana? Näkökulmia bibliografisen datan avaamiseen. Tietolinja 2/2010. Luettavissa osoitteessa <http://urn.fi/URN:NBN:fi-fe201012143105>.

⁴ Tieto käyttöön. Tiekartta tutkimuksen sähköisten tietoaineistojen hyödyntämiseksi. Opetus- ja kulttuuriministeriön julkaisuja 2011:4. Luettavissa osoitteessa http://www.minedu.fi/OPM/Julkaisut/2011/Tiekartta_tutkimuksen_sahkoisten_tietoaineistojen_hyodyntamiseksi.html.

Lisäksi on hyvä huomata, että OKM:lle raportoitavien ei-tieteellisten julkaisujen osalta ei ole mahdollista päästä täydelliseen kattavuuteen, eikä ei-tieteellisistä julkaisuista ole yleensä mahdollista tallentaa affiliaatiotietoja, koska niitä ei ole tapana mainita julkaisun yhteydessä. Kansallisen auktoriteettitietokannan käyttöönotto saattaa helpottaa tilannetta tältä osin jossain määrin, mutta sen sisältämien tietojen hyödyntäminen täysimittaisesti onnistuisi vasta sitten, jos ne voidaan kytkeä yhteen kansallisen tutkijarekisterin kanssa. Niinpä populaarijulkaisujaan raportoivat tutkijat joutuisivat monissa tapauksissa edelleen joko tallentamaan julkaisunsa tiedot kokonaan itse tai ainakin tunnistamaan tietokannasta haravoidun julkaisun omakseen ennen kuin se voidaan liittää mukaan julkaisurekisteriin.

3.3. Kustantajayhteistyö uutena mahdollisuutena

Vaikka Arton tiedontuotanto on tähän asti keskittynyt kirjastoihin, vapaan käyttöoikeuden myötä tiedontuottajien piiriä on aiempaa helpompaa laajentaa myös niiden ulkopuolelle. Tämä edellyttää sitä, että Arto nähdään yhden yksittäisen tietokannan sijasta pikemminkin yhteisenä vapaasti hyödynnettävänä tietovarantona, jonka keskeinen ilmentymä tällä hetkellä on nykymuotoinen Arto-tietokanta.

Kansallisen julkaisurekisterin kannalta merkittävin potentiaalinen uusi tiedontuottajaryhmä ovat tieteelliset kustantajat ja tieteellisten lehtien toimittajat. Kustantajien saaminen mukaan tarjoaisi mahdollisuuksia aivan uudenlaisen julkaisutietojen tuotantoa ja välittämistä palvelevan kansallisen kokonaisjärjestelmän kehittämiseen. Tässä mallissa kustantajat, kirjastot ja tutkimusorganisaatiot pyrkisivät eri intressien välisen vastakkainasettelun sijasta rakentamaan yhteisiä käytäntöjä ja yhteistä palvelua, jotka hyödyttävät yhtä lailla kaikkia osapuolia.

Kustantajayhteistyö tarjoaa mahdollisuuksia Arto-tietokannan tämänhetkisten puutteiden korjaamiseen. Sen avulla voidaan parantaa Arton kattavuutta (ts. saada mukaan uusia lehtiä) ja myös tietojen tallennuksen ajantasaisuutta. Kirjastojen tekemä tallennustyö on riippuvaista painettujen julkaisujen saamisesta luetteloitavaksi, kun taas kustantajalla on vastaavat tiedot käytettävissään parhaassa tapauksessa jo ennen kuin painettu lehti on lähetetty tilaajille. Lisäksi kustantajilla on mahdollisuudet kerätä julkaisuprosessin yhteydessä kansallisen julkaisurekisterin kannalta oleellisia tietoja suoraan julkaisujen tekijöiltä.

Tietosisältöjen osalta kustantajilla on kirjastoja paremmat edellytykset tallentaa kunkin artikkelin julkaisutyyppiä koskevia tietoja ja tekijöiden affiliaatiotietoja. Kustantajayhteistyön kautta olisi mahdollista päästä nykyistä kattavammin käsiksi myös artikkelikokoelmiin ja niiden sisältämiin artikkeleihin, jotka ovat etenkin humanistis-yhteiskunnallisilla tieteenaloilla merkittävä julkaisumuoto. Artikkelikokoelmien, sarjajulkaisujen ja monografioiden tietojen kerääminen edellyttää sitä, että yhteistyöhön otetaan mukaan lehtikustantajien lisäksi myös kirjankustantajat.

Keskeinen haaste kustantajayhteistyön käynnistämisessä on kustantajien motivointi. Millä argumenteilla voimme perustella usein hyvin pienin resurssein toimivalle tieteelliselle kustantajalle sen, että hänen kannattaa osallistua julkaisutietojen tallennukseen? Motivointi edellyttää joka tapauksessa sitä, ettei tietojen tallentamisesta aiheudu kustantajalle kohtuutonta vaivaa, eli tietojen tallennuksen pitää olla helppoa eikä tallennettavia tietoja saa olla liian paljon. Tallennettavan tietomäärän osalta tilannetta helpottaa se, että useimmat kotimaiset tiedelehdet ovat melko pieniä, eli julkaistavien artikkelien määrä vuositasona on yksittäisen lehden kohdalla yleensä joitakin kymmeniä tai korkeintaan sadan tai kahden sadan artikkelin suuruusluokkaa. Tämä tarkoittaa sitä, ettei tallennuksesta aiheudu useimpien lehtien kohdalla kovin suurta työpanosta.

Ilmeisin tietojen tallennuksen kustantajille tarjoama etu on se, että kansallisen järjestelmän kautta julkaisujen tiedot on mahdollista saada nopeasti ja kattavasti tiedonhakijoiden ja potentiaalisten lukijoiden saataville. Kansallisen julkaisurekisterin lisäksi julkaisutiedot päätyvät tiedonhakijoiden ulottuville välittömästi sekä Arto-tietokannan, Nelli-portaalin, rakenteilla olevaan Kansallisen digitaalisen kirjaston asiakasliitty-

män ja tulevaisuudessa mahdollisesti muidenkin vapaasti käytettävää artikkelitietovarantoa hyödyntävien tiedonhakupalveluiden kautta. Tämä parantaa oleellisesti julkaisujen löydettävyyttä ja lisää niiden näkyvyyttä, mikä on kustantajien kannalta merkittävää.

Löydettävyyteen ja näkyvyyteen liittyvien etujen lisäksi kansallista järjestelmää on mahdollista perustella myös sillä, että julkaisutiedot voivat olla sen kautta myös kustantajien itsensä käytettävissä. Järjestelmään syötetyt tiedot on mahdollista saada sieltä ulos erilaisissa formaateissa ja niitä on mahdollista hyödyntää esim. erilaisten indeksien tai verkkosivujen laatimisessa.

Kustantajien motivoiminen ja sitouttaminen on tärkeää myös palvelun ajantasaisuuden ja tallennettavien tietojen laadun kannalta. Yhteistyön kautta saatavien tietojen laatua on vaikea ennakoida, ja voi olettaa että se vaihtelee ainakin jonkin verran niiltä osin kuin tiedot eivät tule järjestelmään valmiiksi tallennetuista tiedoista (esim. lehden nimi) tai syöttölomakkeen valikoista. Aiemmat kokemukset julkaisutietojen keräämisestä verkkolomakkeilla viittaavat kuitenkin siihen, että motivoituneet ja hyvin ohjeistetut tallentajat pystyvät ilman kirjastoalan koulutustakin tuottamaan tietoja, jotka ovat todennäköisesti laadultaan vähintään riittävää tasoa.

Yhteistyösuhteita rakennettaessa on syytä kartoittaa huolellisesti mitkä ovat kunkin kustantajan valmiudet tietojen toimittamiseen ja miten perusteellista työtä nämä ovat valmiita tekemään esim. affiliaatitietojen tallentamiseksi. Toisaalta esim. asiasanoitus ei ole julkaisurekisterien kannalta välttämätöntä, eli se on mahdollista tehdä myös kirjastojen voimin jälkikäteen. Vaikka päämääränä on se, ettei tallennuksesta aiheutuisi kustantajille kohtuutonta vaivaa, on syytä miettiä myös sitä, tarvitaanko sen tueksi jonkinlaista sopimuskäytäntöä, jonka avulla määriteltäisiin osapuolten velvollisuudet ja sovittaisiin järjestelmään tallennetun datan omistusoikeudesta.

3.4. Ulkoisen syöttöjärjestelmän kehittäminen

Kustantajayhteistyö edellyttää uuden tietojen syöttöön ja syöttöprosessin hallintaan soveltuvan tietojärjestelmän rakentamista. Kustantajilla ei näet pääsääntöisesti ole sellaisia omia tietojärjestelmiä, joissa tiedot olisivat valmiina olemassa tai joista niiden siirtäminen olisi mahdollista. Alustavan suunnitelman mukaan Kansalliskirjasto voisi rakentaa artikkelitiedon keruuta varten Arto-tietokannan ulkopuolella toimivan syöttöjärjestelmän, joka mahdollistaisi julkaisutietojen syötön ja tallentamisen Arton ja julkaisurekisterin kannalta yhteensopivassa formaatissa. Samaa järjestelmää olisi mahdollista käyttää myös muita julkaisutyyppisiä koskevan tiedon keruuseen, vaikka sen hyödyt ovat suurimmat nimenomaan lehtiartikkelien kaltaisen paljon toistuvia tietoja sisältävän aineiston kohdalla.

Tällä hetkellä Arto-tietokannan tallennustyö tapahtuu Voyager-kirjastojärjestelmän luettelointiklientillä, joka ei sovellu kirjastojen ulkopuoliseen käyttöön. Artikkelitietojen syöttöön soveltuvan helppokäyttöisen verkkolomakkeen rakentaminen näyttäisi kuitenkin alustavien selvitysten perusteella olevan mahdollista. Järjestelmä voisi toimia itsenäisesti Arto-tietokannan ulkopuolella ja tietojen siirto järjestelmästä toiseen tapahtuisi teknisten rajapintojen kautta. Suurimmat tekniset haasteet liittyvät tallennusprosessien suunnitteluun sekä niiden hallintaan liittyvien toiminnallisuuksien rakentamiseen. Lisäksi syöttölomakkeen kytkeminen kirjastotietokantaan on jossain määrin haasteellista Voyagerin puutteellisten rajapintojen vuoksi. Kansalliskirjasto on juuri selvittämässä mahdollisuuksia rakentaa vastaavia kirjastojärjestelmän ulkopuolella toimivia tietojen syöttöön ja tietokannan jo sisältämien tietojen hallintaan tarkoitettuja työkaluja, ja tätä kehitystyötä voitaisiin todennäköisesti hyödyntää myös tässä hankkeessa.

Syöttöjärjestelmä toimii siten, että tallentajat kirjautuvat järjestelmään omilla tunnuksillaan ja syöttävät tyypillisesti kerralla oman lehensä tuoreimman numeron artikkelien tiedot. Järjestelmä sisältää valmiiksi kunkin julkaisun perustiedot (esim. lehden nimi, ISSN, julkaisijan tiedot) ja tukee tallennusprosessia tarjoamalla toistuvat tiedot (esim. julkaisutyyppi, affiliaatitiedoissa toistuvat organisaatiot) standardoidussa

muodossa valmiina valikoina tai siten, että järjestelmä osaa ehdottaa niitä ensimmäisten kirjainten perusteella. Oletuksena on, että ainakaan alkuvaiheessa ei ole käytettävissä kansallista tutkijarekisteriä, josta olisi mahdollista saada myös kirjoittajien tiedot, joten ainakin kirjoittajien nimet tiedot täytyy syöttää käsin. Syötetyt tiedot tallentuvat soveltuvassa muodossa välivarastoon syöttöjärjestelmään, josta ne tarkistuksen ja mahdollisen täydennyksen jälkeen päätyvät joko välittömästi tai säännöllisin väliajoin tehtävinä eräajoin Arto-tietokantaan.

Syöttöjärjestelmän oman tallennuslomakkeen lisäksi voi olla perusteltua rakentaa myös mahdollisuus tuoda tietoja järjestelmään muista ulkoisista tietojärjestelmistä tai julkaisutietoja sisältävistä tiedostoista (esim. Excel-tiedostot tai erilaisista viitteidenhallintajärjestelmistä tulostetut tiedostot). Tämä edellyttää luonnollisesti sitä, että importoitavat tiedot ovat jossain standardoidussa formaatissa ja sisältävät Arto-tietokannan ja kansallisen julkaisurekisterin kannalta keskeiset tietoelementit. Tätä mahdollisuutta voidaan hyödyntää esim. siinä tapauksessa jos kustantajalla on oma tietojärjestelmä, josta on mahdollista saada joko kaikki Arton/Jure:n tarvitsemat tiedot tai osa niistä. Toinen esimerkitapaus on Tieteellisten seurain valtuuskunnan ylläpitämä Tieteelliset verkkolehdet -palvelu (<http://ojs.tsv.fi/>), joka sisältää tällä hetkellä kymmenkunnan lehden aineistoja. Siinä tapauksessa että alkuperäinen tietojärjestelmä ei sisällä kaikkia oleellisia tietoja, syöttöjärjestelmä tarjoaa mahdollisuuden täydentää tietoja jälkikäteen.

3.5. Kustantajayhteistyön ja uusien työvälineiden vaikutus kirjastoissa tehtävään työhön

Vaikka tieteellisten julkaisujen tietoja saataisiin ennen pitkää laajassa mittakaavassa kustantajilta, tämä ei kuitenkaan tarkoita sitä, etteikö myös kirjastojen työpanosta ja sitoutumista yhteisen palvelun tuottamiseen tarvittaisi edelleen. Pikemminkin kustantajilta saatu perusdata mahdollistaa sen, että tallennuksen prosesseja voidaan kehittää ja kirjastojen työpanosta voidaan joiltakin osin suunnata uudelleen.

Kirjastojen panosta tarvitaan luonnollisesti edelleen sellaisten julkaisujen luetteloinnissa, joiden tietoja ei saada kustantajilta. On epärealistista odottaa, että kaikki kustantajat ryhtyvät yhdessä yössä tallentamaan tietoja kansallisen artikkelitietovarannon tarpeisiin. Kirjastot voivat myös täydentää ja rikastaa kustantajilta saatuja tietoja (mm. asiasanoitus). Muita kirjastoille soveltuvia tehtäviä voivat olla tallennettujen tietojen tarkistaminen ja hyväksyminen ennen kuin ne siirretään syöttöjärjestelmästä Arto-tietokantaan. Lisäksi tallennuksen koordinoimista ja siihen liittyvää neuvontaa on mahdollista hajauttaa jakamalla kirjastojen kesken vastuualueita ja vastuulehtiä esim. tieteenaloitain.

Kustantajien tekemän tallennustyön lisäksi sitä varten kehitettävää syöttöjärjestelmää olisi mahdollista hyödyntää myös muussa Voyager-kirjastojen ulkopuolella tehtävässä tallennustyössä. Jos lomake osoittautuu oleellisesti 1990-luvun teknologiaa edustavaa Voyagerin luettelointiklienttiä tehokkaammaksi tallennusvälineeksi, mikään ei tietysti estä sen käyttämistä myös Voyager-kirjastoissa. Syöttöjärjestelmälle on lisäksi mahdollista kehitellä monenlaisia muita käyttötarkoituksia, jotka eivät suoraan liity Jure-projektin päämääriin (ks. LIITE 2).

4. Miten eteenpäin?

Kuten edellä on käynyt ilmi, Kansalliskirjaston ylläpitämiä viitetietokantoja (etenkin Artoa) on mahdollista kehittää siten, että ne soveltuvat nykyistä paremmin kansallisen julkaisurekisterin tietolähteeksi. Aikataulullisesti tämä on selkeästi Jure II:een kytkeytyvä hanke, sillä Jure I -vaiheessa tietokannoista saatavien viitetietojen tuottama hyöty on vielä suhteellisen vähäinen. Tämä antaa suhteellisen väljän aikataulun tarvittavien teknisten ratkaisujen kehittämiseen sekä hankkeen edellyttämien yhteistyömallien ja prosessien luomiseen.

Tietokantojen tietosisältöjä on mahdollista laajentaa jossain määrin myös nykyisen kirjastojen tekemään luettelointiin pohjautuvan toimintamallin puitteissa. Vaikka kansallinen yhteisluettelo ja muut ajankohtaiset kehittämishankkeet uudistavat muutenkin kirjastojen välistä työnjakoa ja lisäävät kansallista yhteistyötä, kirjastojen sitouttaminen Arto-tietokannan luettelointiin käytettävien resurssien merkittävään lisäämiseen voi kuitenkin olla vaikeaa. Tämä huolimatta siitä, että se tarjoaisi palkintona nykyistä laadukkaamman yhteisessä käytössä olevan kansallisen palvelun ja tietovarannon, ja säästäisi samalla tätä kautta OKM:n tiedonkeruuta varten tehtävään julkaisutietojen tallennukseen paikallisesti käytettäviä resursseja.

Vaikka kustantajayhteistyön kehittämiseen liittyy etukäteen monia epävarmuustekijöitä, se avaisi kuitenkin kokonaan uudenlaisia mahdollisuuksia julkaisutietojen tallentamista ja välittämistä palvelevan kansallisen kokonaisjärjestelmän luomiseen suhteellisen vähäisillä kustannuksilla. Kehittämishankkeen toteuttaminen edellyttää sekä uuden tietojärjestelmän rakentamista että yhteisen tahtotilan löytämistä toisaalta tieteellisten kirjastojen ja toisaalta tieteellisten kustantajien kesken. Kirjastojen osalta tarvitaan keskustelua Arto-tallennuksen kehittämisen prioriteeteista, tallennukseen käytettävistä resursseista ja työnjaon uudistamisen tarjoamista mahdollisuuksista. Kustantajien kannalta tietojen tallennus on kokonaan uusi kuvio, eli sen lanseeraaminen vaatii runsaasti pohjustusta ja laajaa keskustelua.

Jos kehittämishanketta päätetään lähteä toteuttamaan, alkuvaiheessa on todennäköisesti järkevintä lähteä liikkeelle pilottiprojektin kautta, jotta tietojen tallennuksesta ja syöttöjärjestelmän toiminnasta saadaan riittävästi kokemuksia ennen kuin järjestelmä otetaan käyttöön lopullisessa muodossa. Tätä selvitystä varten on toistaiseksi tehty vain hyvin alustavia yhteydenottoja tiedekustantajien suuntaan, eli varsinaisen kehittämishankkeen valmistelu edellyttäisi joka tapauksessa vielä laajempaa selvitystä kustantajien valmiuksista ja kiinnostuksesta osallistua julkaisutietojen tuotantoon. Kirjastosektorien ulkopuolella keskeisinä yhteistyökumppaneina näyttäytyvät etenkin Tieteellisten seurain valtuuskunta ja Suomen tiedekustantajien liitto, joilla on valmiit yhteysverkot kotimaisiin tieteellisiin kustantajiin.

Jos teknisen järjestelmän kehitystyö ja kustantajayhteistyön valmistelu alkaisi syksyllä 2011, palvelun pilotointi olisi mahdollista aloittaa vuonna 2012 ja laajempi käyttöönotto todennäköisesti vuoden 2013 kuluessa. Tämä edellyttää sitä, että toiminnan koordinointiin ja suunnitteluun, tekniseen kehitystyöhön sekä kirjastojen ja kustantajien kanssa käytäviin neuvotteluihin olisi käytettävissä riittävät resurssit. Vaikka myös teknisten järjestelmien rakentaminen vaatii työtä, on selvää, että hankkeen suurimmat haasteet liittyvät markkinointiin ja tiedotukseen, eri toimijoiden yhteisen tahtotilan löytämiseen sekä tallennuksen prosessien rakentamiseen ja vakiinnuttamiseen.

5. Liitteet

LIITE 1: Muut kotimaiset artikkelitietokannat

LIITE 2: Mahdollisuuksia syöttöjärjestelmän jatkokehitykselle

LIITE 1: Muut kotimaiset artikkelitietokannat

Artoa perustettaessa vuonna 1995 osa kotimaisista artikkelitietokantojen tuottajista jäi sen ulkopuolelle. Osa oli ollut mukana Artoa edeltäneessä Kati-yhteistyössä, mutta Arton toimintamalli ei kuitenkaan tässä vaiheessa sopinut niille joko toiminnan rahoitukseen tai tietosisältöjen muotoon liittyvistä syistä. Tilanne on osoittautunut pysyväksi, ja niinpä Arton lisäksi Suomessa tuotetaan tällä hetkellä seuraavia artikkeliviitteitä sisältäviä tietokantoja:

- BTJ Aleksi (<http://aleksi.btj.fi/>)
Kotimaisten aikakaus- ja sanomalehtien artikkeliviitteitä, 685.000 viitettä, 35.000/vuosi, maksullinen.
- Tali - artikkeliviitteitä suomenkielisiin tekniikan alan lehtiin (<http://www.otalib.fi/tkk/tali/>)
Aalto/TKK, 330.000 viitettä, maksullinen, myydään TKK:n tuottamien Tenttu-tietokantojen yhteydessä.
- FINP - Finnish Periodicals Index in Economics and Business
(<http://helecon.lib.hse.fi/FI/database/FINP/>)
Aalto/HKKK, 190.000 viitettä, 9.000/vuosi, maksullinen, myydään joko erillisenä tai muiden Helecon-tietokantojen kanssa paketoituna.
- Medic - kotimainen terveystieteiden viitetietokanta (<http://www.terkko.helsinki.fi/medic/>)
HY/Terkko, 100.000 viitettä, 4.000/vuosi, maksullinen.
- Leo (<http://pretty.ttl.fi/LEO/search.htm>)
Työsuojeluun liittyviä lehtiartikkeli- ja säädösviitteitä vuosilta 1978-2008, Työterveyslaitos, 35.000 viitettä, vapaassa käytössä.
- Press (<http://press.bragespressarkiv.fi/>)
Ruotsinkielisten sanomalehtien artikkeliviitteitä, Brages pressarkiv, 250.000 viitettä, vapaassa käytössä.

Vaikka artikkelitietokantoja on monta, ne ovat kuitenkin erikoistuneet erilaisiin aineistoihin ja täydentävät toisiaan. Tietokantojen välillä on jonkin verran päällekkäisyyksiä, mutta esim. Arton ja Aleksin osalta näitä on pyritty yhteistyössä vähentämään siten, että tietyt lehdet luetteloidaan vain jompaankumpaan tietokantaan. Artoon verrattuna Aleksi on muutenkin keskittynyt selkeästi suurta yleisöä kiinnostaviin aikakauslehtiin ja sanomalehtiin, vaikka siihen luetteloidaan myös jonkin verran tieteellisiä lehtiä.

Suurempi ongelma onkin se, että kaikkien tietokantojen tuotantoprosessit ja tekniset ratkaisut ovat erillisiä ja erilaisia, eikä niistä muodostu yhtenäistä kokonaisuutta. Tähän liittyy myös se, että tietokantojen sisältämät tiedot eivät välttämättä ole Arton käyttämässä Marc-formaatissa. Lisäksi maksulliset tietokannat on hankittu vain osaan tutkimusorganisaatioista, eli niiden hyödyntämiseen kansallisen julkaisurekisterin kaltaisessa palvelussa voi liittyä sopimusteknisiä esteitä.

Kotimaisten artikkelitietokantojen edelleen jatkuva hajaannus periytyy siis viidentoista vuoden takaa, ja tämän päivän näkökulmasta katsottuna on ilmeistä, että toiminnan voisi ainakin korkeakoulujen ylläpitämien tietokantojen osalta organisoida toisinkin, olettaen että yhteistyön periaatteista löydetään yhteisymmärrys. BTJ:n tuottama Aleksi on tässä suhteessa hankalampi tapaus, sillä tietokannan toiminta perustuu kokonaan lisensioinnista saatuihin tuloihin ja sen keskeinen asiakasryhmä (yleiset kirjastot) on korkeakoulusektorin ulkopuolella.

Vaikka Arto ja Fennica/Linda kattavat melko hyvin erityisesti humanistis-yhteiskuntatieteellisten alojen kotimaiset julkaisut, kansallisen julkaisurekisterin näkökulmasta olisi mielekästä, jos myös muiden alojen kotimaisia artikkeliviitteitä olisi mahdollista saada mahdollisimman kattavasti valmiina julkaisurekisterin käyttöön. Tämä edellyttää keskustelua nykyisten tietokantojen ylläpitäjien kanssa siitä, miten paljon kansallisen julkaisurekisterin kannalta relevanttia aineistoa ne sisältävät ja millä ehdoilla esim. tekniikan tai kauppatieteiden alan artikkeliviitteitä voitaisiin hyödyntää Kansallisessa julkaisurekisterissä. Keskeinen kysymys on, miten helppoa tietokantoja olisi käyttää nykymuodossaan julkaisurekisterin tietolähteinä, vai olisiko ne järkevämpää koota Arton yhteyteen yhtenäiseksi, vapaasti hyödynnettäväksi tietovarannoksi?

Erillisten tietokantojen hyödyntäminen edellyttäisi sitä, että niihin kaikkiin rakennettaisiin tarvittavat tekniset rajapinnat ja niiden tietosisältöä kehitettäisiin vastaamaan kansallisen julkaisurekisterin tarpeita. Tietokantojen yhdistäminen osaksi yhteistä tietovarantoa taas edellyttäisi sopimusta toimintaan liittyvästä vastuiden jaosta ja mahdollisesti myös jonkinasteisia taloudellisia järjestelyjä. Toisaalta yhtenäisten prosessien luominen ja yhteisen tietojärjestelmän käyttöönotto tarjoaisi mahdollisuuksia sekä toiminnan tehostamiseen että kustannussäästöihin, ja Arton yhteydessä toteutuva vapaa saatavuus laajentaisi myös merkittävästi viitetietojen käyttöä ja parantaisi tiedonhakijoille tarjottavaa palvelua.

LIITE 2: Mahdollisuuksia syöttöjärjestelmän jatkokehitykselle

Kansalliseen julkaisurekisteriin liittyvien päämäärien lisäksi selvityksessä esitettyä julkaisutietojen syöttöjärjestelmää ja sen avulla syntyvää tietovarantoa olisi lisäksi mahdollista kehittää edelleen siten, että niiden pohjalta olisi mahdollista rakentaa muunlaisiin tarpeisiin liittyviä prosesseja ja palveluita.

1. Kokotekstitiedostojen tallentaminen syöttöjärjestelmän kautta

Julkaisujen viitetietojen keruuseen tarkoitettua syöttöjärjestelmää olisi mahdollista kehittää myös siten, että sen kautta voisi tallentaa samalla kertaa viitetietojen lisäksi myös kokotekstitiedostoja. Teknisesti tämän toiminnon toteuttaminen olisi melko yksinkertaista, ja yhdistettynä kuvailutietoihin tiedostoja olisi mahdollista hyödyntää monenlaisissa yhteyksissä. Tiedostojen formaattiin ja käsittelyyn liittyy tosin myös joitakin haasteita: esim. tallennetaanko aineistot artikkeli kerrallaan, kokonaisina numeroina vai sekä että?

Mitä tahansa tätä kautta saadulla tiedostoilla tehdäänkin, siihen pitää olla oikeudenhaltijoiden lupa. Tiedostojen käyttöoikeuksista pitää sopia tallennuksen yhteydessä kustantajien kanssa (olettaen, että kustantaja on hankkinut julkaisujen tekijöiltä riittävät oikeudet). Syöttöjärjestelmän sisällä sinne tallennetut tiedostot ovat lähtökohtaisesti vain hyvin rajatusti lehden itsensä sekä palvelun ylläpitäjien ja muiden auktorisoitujen henkilöiden (esim. kuvailutietojen tarkistamista ja rikastamista tekevä luetteloija) saatavilla. Lisäksi syöttöjärjestelmä olisi mahdollista kytkeä Kansallisen digitaalisen kirjaston puitteissa suunniteltavaan digitaalisten aineistojen pitkäaikaissäilytysjärjestelmään siten, että sitä voidaan käyttää julkaisujen luovutukseen tiedostomuodossa.

Jos kustantajan kanssa tehty sopimus sen sallii, julkaisut on mahdollista välittää syöttöjärjestelmästä esim. Elektran kaltaiseen maksulliseen julkaisupalveluun tai vaihtoehtoisesti open access -julkaisuina vapaasti verkkoon. Jos kustantaja sallii julkaisun rinnakkaistallennuksen kustantajan versiota käyttäen, artikkeli-tiedosto on mahdollista välittää myös asiaankuuluvaan julkaisuarkistoon siten, että se tulee saataville lehden määrittämän embargo-ajan jälkeen. Tiedostojen syötön yhteydessä niille on mahdollista antaa myös pysyvät tunnisteet ja tunnisteisiin perustuvat pysyvät verkko-osoitteet, jotka välittyvät eteenpäin sekä mahdolliseen julkaisupalveluun että julkaisurekistereihin.

2. Syöttöjärjestelmän käyttö takautuvien aineistojen viitetietojen syöttämiseen

Uusien julkaisujen viitetietojen lisäksi syöttöjärjestelmän olisi mahdollista syöttää takautuvasti vanhojen julkaisujen viitetietoja. Järjestelmän kautta syötettyjä takautuvien aineistojen kuvailutietoja on mahdollista hyödyntää myös digitoitujen aineistojen julkaisemisessa.

3. Syöttöjärjestelmän laajentaminen avoimeksi sosiaalseksi verkkopalveluksi

Syöttöjärjestelmää voisi periaatteessa kehittää myös laajemmalle yleisölle suunnatuksi sosiaalseksi verkkopalveluksi, jonne kuka tahansa voi tallentaa erilaisten lehtien ja muidenkin julkaisujen viitetietietoja. Verkosta löytyy jo nyt useita yksityishenkilöiden verkkosivuja, joille on koottu esim. harrastelehtien sisällysluetteloita, ja tätä samaa aktiivisuutta olisi kenties mahdollista kanavoida esim. LibraryThing-palvelun (<http://www.librarything.com/>), Wikipedian (http://fi.wikipedia.org/wiki/Main_Page) tai Gutenberg-projektin (http://www.gutenberg.org/wiki/Main_Page) malliin yhteisen palvelun tuottamiseen. Jotta tallentajien motivaatio ja tietojen laatu saataisiin optimoitua, palveluun pitäisi rakentaa yhteisöllistä päätöksentekoa ja sosiaalista kontrollia tukevia toiminnallisuuksia.